Ratings of Decision-Making Attributes In a Junior Leader Course

Scott A. Beal

**United States Army Research Institute
for the Behavioral and Social Sciences**

**Infantry Forces Research Unit
Scott E. Graham, Chief**

October 2002

**United States Army Research Institute
for the Behavioral and Social Sciences**

20021126 104

**U.S. Army Research Institute
for the Behavioral and Social Sciences**

A Directorate of the U.S. Total Army Personnel Command

**ZITA M. SIMUTIS
Acting Director**

Technical Review by
Harnam Singh

NOTICES

REPORT DOCUMENTATION PAGE

1. REPORT DATE October 2002	2. REPORT TYPE Final	3. DATES COVERED (from. . . to) January-August, 2002

4. TITLE AND SUBTITLE Ratings of Decision-Making Attributes in a Junior Leader Course	5a. CONTRACT OR GRANT NUMBER
	5b. PROGRAM ELEMENT NUMBER 62785

6. AUTHOR(S) Scott A. Beal (U.S. Army Research Institute)	5c. PROJECT NUMBER A790
	5d. TASK NUMBER 204
	5e. WORK UNIT NUMBER H01

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) U.S. Army Research Institute for the Behavioral and Social Sciences Infantry Forces Research Unit P. O. Box 52086 Fort Benning, GA 31995-2086	8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES) U.S. Army Research Institute for the Behavioral and Social Sciences 5001 Eisenhower Avenue Alexandria, VA 22333-5600	10. MONITOR ACRONYM ARI
	11. MONITOR REPORT NUMBER Research Note 2003-02

12. DISTRIBUTION/AVAILABILITY STATEMENT

Approved for public release; distribution is unlimited.

13. SUPPLEMENTARY NOTES

14. ABSTRACT *(Maximum 200 words)*: Cadre and students provided ratings of decision-making attributes during two junior leader courses. Results showed that cadre ratings of students' attributes improved over time, whereas student self-ratings did not show improvements. Initial student ratings tended to be inflated. Cadre ratings were more conservative than student ratings regardless of rating time. Research on naturalistic decision-making and self-appraisal were considered for the purposes of interpreting and explaining discrepancies across rating sources. It was suggested that discrepancies arose because of differences in raters' experience, knowledge, and because of inconsistent or ambiguous performance comparison standards.

15. SUBJECT TERMS Decision-making, Self-Report, Self-Ratings, Self-Appraisal, Rating Discrepancies, Naturalistic Performance Comparison Standards

SECURITY CLASSIFICATION OF			19. LIMITATION OF ABSTRACT	20.NUMBER OF PAGES	21. RESPONSIBLE PERSON (Name and Telephone Number)
16. REPORT Unclassified	17. ABSTRACT Unclassified	18. THIS PAGE Unclassified	Unlimited	22	Scott A. Beal (706) 545-2356

RATINGS OF DECISION-MAKING ATTRIBUTES IN A JUNIOR LEADER COURSE

CONTENTS

List of Tables

Ratings of Decision-Making Attributes in a Junior Leader Course

Introduction

One of the challenges facing the twenty-first century Army is producing competent leaders. The Army's success during wartime operations is, in part, a reflection of its ability to train leaders effectively. Readiness training conducted to high standards is an essential obligation for Army leaders during peace and war. It is no surprise, then, that the Army places a premium on building competent leaders who can influence soldiers serving under their guardianship.

Army leaders are often called upon to execute missions in rapidly changing environments. Competent and successful leaders minimize loss by making decisions that lead to positive outcomes. Effective decision-making by leaders under conditions of uncertainty has been a focus of the Army's attention (dt ogilvie & Fabian, 1999). Researchers are challenged with the task of determining whether existing training methods improve leadership and decision-making, which in turn improve the overall effectiveness of Army operations. In making these types of determinations, researchers often rely on self-reports from trainers and trainees (see Farh, Werbal, & Bedeian, 1988).

The accuracy of data collected by self-report methods has been a point of contention for some researchers (Critchfield, 1993; Critchfield, Tucker & Vuchinich, 1998). Behavioral observations of actual leader performance combined with self-reports are an important contribution to leader assessment. However, there are occasions when self-reported perceptions of performance contradict actual performance, such as when individuals who perform under time constraints overestimate or underestimate their performance (Critchfield, 1996). A challenge for researchers is to determine when and at what magnitude self-report inaccuracies occur.

When researchers measure leader effectiveness, it is conventional to employ the use of paper and pencil surveys that allow individuals to rate themselves on leader attributes, traits, behaviors, or competencies. In the military literature, attempts to validate self-ratings or to quantify discrepancies between ratings and actual performance are the exception. Tremble (1992) compared ratings of leader competency attributes from multiple rating sources and ratings of unit performance relative to expected standards. During observations of unit training, observer/controllers rated the competency attributes of platoon leaders and platoon sergeants and the performance of the units under their command. Platoon leaders were also rated by platoon sergeants and by squad leaders, and platoon leaders, squad leaders, and squad members rated platoon sergeants. Results showed that observer/controller mean ratings were the most conservative, squad leader and squad member mean ratings were slightly higher, and platoon leader and platoon sergeant mean ratings of each other were the highest. Although no self-ratings were collected and assessed, upon comparison, ratings of leader competency attributes from observer/controllers correlated highly with observer/controller ratings of unit performance, which suggested

that ratings from experienced raters can be accurate representations of actual performance.

Studies of two junior leader courses were conducted recently at Fort Benning, GA. Researchers developed a self-report metric that allowed students to rate themselves over time on leadership attributes. Cadre also rated students over time on the same attributes. The purpose of these assessments was not to corroborate self-reports with performance measures directly. However, discrepancies between self-ratings and cadre ratings of student decision-making attributes that arose during the junior leader courses provided an impetus for reconciling ratings differences. The present study compared cadre ratings and self-ratings on the leadership attributes related to decision-making.

Methods used to rate leadership attributes were identical for both junior leader courses. However, information from junior leader course 1 was provided to cadre who rated students in junior leader course 2. Because this information was provided, and because the students and cadre who participated in junior leader course 2 were different from those who participated in course 1, data from each course were analyzed and reported separately.

<div align="center">Junior Leader Course 1</div>

Method

One hundred seventy-two students received a leadership attributes survey during the third and sixth weeks of the course. The instruction sheet included a seven-point scale that participants were instructed to use to rate each survey item (1 = much lower than average, 2 = lower than average, 3 = slightly lower than average, 4 = average, 5 = slightly higher than average, 6 = higher than average, 7 = much higher than average, CR = cannot rate) (see Appendix A).

Cadre were given a survey for each student in their respective platoons during the third and sixth weeks of course training and instruction. This survey was identical to the survey used by self-raters, except that cadre were instructed to rate students' performance.

According to the Field Manual (FM) 100-5 *Operations* (Department of the Army [DA], 1993) and the FM 3-0 *Operations* (DA, 2001), effective decision making combines judgment with information as an element of combat power: it requires knowing if to decide, when to decide, and what to decide.

The survey, along with the above definition of decision-making, was given to four subject-matter experts (SME). SMEs were chosen based on their military experience. Three SMEs were retired Infantry officers and another was a retired senior Infantry NCOs. The SMEs were instructed to choose items from the survey that were relevant to decision-making and relevant to assessing whether the decision-making attributes of students improved over time. A subset of nine survey items was chosen by at least

<div align="center">2</div>

three of the SMEs and was included in the analyses (see Appendix A). The decision-making subset included the following items: Attention to Detail; Creativity; Initiative, Self-Motivation, Persistence; Knowledge of Military Tactics/Expertise; Listening; Seek Innovative Solutions; Situational Awareness, Perception; Sound Judgment, Logical Reasoning; Time Management.

Results

Self-ratings at times 1 and 2 were distributed similarly across the rating categories (see Table 1). Seventy-five percent of the leaders rated themselves above average, with approximately 13% placing themselves in the highest category of "much higher than average." Only about 18% rated themselves as average. Given the high self-appraisals reflected in these ratings, if performance improvements really occurred during the course, it would have been difficult for the improvements to be represented by subsequent ratings.

Table 1
Percentage of Students in Each Rating Category for Junior Leader Course 1

Category	Self-Rating 1	Self-Rating 2	Cadre Rating 1	Cadre Rating 2
Much Higher	13%	12%	1%	3%
Higher	34%	38%	11%	16%
Slightly Higher	28%	29%	24%	34%
Average	19%	17%	52%	39%
Slightly Lower	4%	4%	10%	5%
Lower/ Much Lower[a]	1%	1%	1%	3%

Note. Cell percentages are the average of the nine decision-making items.
[a]Two rating categories were combined.

Repeated measures analyses of variance (ANOVA) were performed on each of the nine items across self-ratings times 1 and 2 (see Table A-1). The Attention to Detail and Knowledge of Military Tactics/Expertise items significantly increased over time. An index of effect size showed that these increases were small. Further analyses revealed that for Attention to Detail, 38% of the students reported an improvement at time 2, whereas only 19% rated themselves lower. A similar finding applied to Knowledge of Military Tactics/Expertise. For this item, 37% of the students reported an improvement at time 2, whereas 21% rated themselves lower (see Table A-9).

Under conditions where two cadre rated a single student at time 1 on any decision-making item, the time 1 rating from one cadre was summed with the time 1 rating from the other cadre. The sums were then averaged to produce a combined rating. Time 2 ratings were averaged in the same way. If one of the cadre rated a student on any item only once, that rating was eliminated from the analyses. In other words, multiple cadre ratings for a single student at time 1 were combined with ratings at time 2 only when both cadre rated the student at both times. Otherwise, the rating was omitted.

The cadre ratings were more conservative than the self-ratings at both times (see Table 1). Initially, the cadre rated 52% of the students as average. At time 2, this percentage decreased to 39%, with slightly higher percentages occurring for the "slightly higher than average" category. However, the percentage of students rated as "much higher than average" remained low (3% and less). Of interest is that the cadre also showed a tendency to give high ratings, in that the distribution of responses was not symmetric about the "average" category. No more than 11% of the students were rated as below average. Given that the majority of the initial responses by cadre was "average," there was room for improvement.

Repeated measures ANOVAs were performed on each of the nine decision-making items across cadre ratings times 1 and 2. Significant increases over time were found for all items except Listening (see Table A-2). Effect size values for cadre ratings ranged from small to moderate, depending on the item.

Statistical comparisons of self-ratings time 1 with cadre ratings time 1 were conducted using repeated measures ANOVAs (see Table A-3). For example, self-rating time 1 of the item Attention to Detail was compared to cadre rating time 1 of Attention to Detail. ANOVAs were conducted for each item across self- and cadre ratings. Results demonstrated that self-ratings time 1 were significantly higher than cadre ratings time 1 for all nine items. Effect size values showed that differences were moderate to large depending on the item. Similar results were found when self-ratings time 2 were compared to cadre ratings time 2. All items were significantly higher for self-ratings except Knowledge of Military Tactics/Expertise (see Table A-4). Values of effect size showed that differences were moderate to large.

In order to assess the consistency of ratings of decision-making items, self-ratings and cadre ratings for leader course 1 were correlated for each item across time. For example, self-ratings time 1 for the item Attention to Detail was correlated with self-ratings time 2 for the same item. A separate correlation was performed for each item pair for all self-ratings and cadre ratings. Correlations ranged from .36 to .63 for the student self-ratings and from .33 to .60 for the cadre ratings (see Table A-10). All correlations were significant ($p<.05$).

4

Method

The leader survey and an instruction sheet were administered to 155 students during the third and sixth weeks of the junior leader course 2. The survey was given to students under the same conditions as in course 1.

Cadre were given a survey for each student in their respective platoons during the fourth and sixth weeks of the course. This survey was identical to the one used by self-raters. Multiple cadre ratings of a single student were averaged the same way they were averaged in junior leader course 1.

Results

Self-ratings at times 1 and 2 were distributed similarly across rating categories (see Table 2). Seventy-four percent of the students rated themselves above average, with 11% and 7% rating themselves much higher than average for rating times 1 and 2 respectively. Only 21% and 22% rated themselves as average. Because of the high percentage of above average ratings initially, if performance improvements occurred during the course, they would not have been reflected in subsequent ratings.

Table 2
Percentage of Students in Each Rating Category for Junior Leader Course 2

Category	Self-Rating 1	Self-Rating 2	Cadre Rating 1	Cadre Rating 2
Much Higher	11%	7%	3%	2%
Higher	34%	31%	10%	31%
Slightly Higher	29%	34%	23%	33%
Average	21%	22%	44%	24%
Slightly Lower	4%	6%	18%	8%
Lower/ Much Lower[a]	1%	1%	1%	1%

Note. Cell percentages are the average of the nine decision-making items.
[a]Two rating categories were combined.

Repeated measures ANOVAs were performed on each of the nine decision-making items across self-ratings 1 and 2 (see Table A-5). Ratings for Creativity; Initiative, Self-Motivation, Persistence; Listening; and Sound Judgment, Logical Reasoning significantly decreased over time. Effect size values showed that changes were small.

5

For Creativity, 20% of the students reported an improvement at time 2, whereas 37% rated themselves lower. For Initiative, Self-Motivation, Persistence, 18% rated themselves higher at time 2, whereas 37% rated themselves lower. For Listening, 20% reported an improvement and 39% rated themselves lower. For Sound Judgment, Logical Reasoning, 19% rated themselves higher, whereas 34% had lower ratings (see Table A-9).

Cadre ratings of students were more conservative than self-ratings at both times (see Table 2). Initially, cadre rated 44% as "average." At rating time 2, this percentage decreased to 24%, whereas Aslightly higher than average@ ratings increased from 23% to 33%. Ratings from the Ahigher than average@ category increased from 10% to 31%, whereas the percentage of students rated as Amuch higher than average@ remained below 3%. Given that 44% of initial responses by cadre were Aaverage,@ there was room for improvement.

Repeated measures ANOVAs were performed on all nine decision-making items across cadre ratings 1 and 2 (see Table A-6). Significant increases over time occurred for all nine items. Effect sizes ranged from moderate to large values.

Repeated measures ANOVAs were used to compare self- and cadre ratings (see Table A-7). Results demonstrated that self-ratings time 1 were significantly higher than cadre ratings time 1 for all decision-making items except Knowledge of Military Tactics/Expertise. Effect size values were moderate to large. For comparisons of self-ratings time 2 with cadre ratings time 2, self-ratings for four items were significantly higher than cadre ratings. Those four items were Creativity; Situational Awareness, Perception; Listening; and Sound Judgment, Logical Reasoning. Effect size values were relatively small (see Table A-8). All correlations were significant ($p<.05$).

The consistency of decision-making items was assessed by correlating ratings for each item across time. For instance, self-ratings time 1 for the item Attention to Detail was correlated with self-ratings time 2 for the same item. For student self-ratings correlations ranged from .46 to .66. For cadre ratings correlations ranged from .37 to .57 (see Table A-10).

Discussion

One purpose of the present study was to determine whether student ratings and cadre ratings of decision-making attributes increased over time. In general, results suggested that although students did not report improvements over time, cadre did report improvements. Findings also demonstrated that, in general, self-ratings were higher than cadre ratings regardless of rating time. An exception to this finding occurred when junior leader course 2 self-ratings time 2 were compared with cadre ratings time 2 from course 2. Only four of the nine decision-making items were significantly higher for self-ratings. That two independent groups of self-raters failed to report improvements over time and that two groups of cadre did report improvements suggested that the discrepancy might be a function of rating biases. The question of

6

discrepancy might be a function of rating biases. The question of whether students or cadre were responsible for the rating biases needed to be determined.

One explanation of the discrepancy between cadre and students ratings is that the lack of reported improvements by self-raters may have been the result of inflated ratings at time 1. If self-ratings inflations occurred at time 1, then they would have served to restrict the range of potential improvements that could have been reflected in self-ratings at time 2. The distribution of the percentage of responses of rating categories showed that initial self-ratings were inflated.

As a result of inflated self-ratings at time 1, self-ratings at time 2 could not reflect substantial improvements in performance for student decision-making attributes. Thus, the absence of improvements in self-ratings over time may not have been an accurate reflection of improvements in actual performance. It is possible that cadre made more accurate reports of decision-making performance as a result of their experience and more thorough observations of students' actual performance. It is also possible that improvements were reported by cadre to reflect their ability to train junior leaders effectively. However, the absence of data from systematic behavioral observations during the junior leader courses made it impossible to quantify the discrepancy between reports of performance and actual performance. This was one of the major criticisms by Critchfield et al. (1998) regarding the use of self-report data in the absence of objective performance measures in previous studies. Yet, in spite of this criticism and others like it, self-report metrics continue to be used in many organizations (including military training environments) where more intrusive data collection methods are often prohibited (Mabe & West, 1982).

In an attempt to interpret and explain discrepancies between cadre ratings and self-ratings that emerged during the junior leader courses, prior research on decision-making and on self-appraisal was considered. Klein (1993) proposed a "naturalistic" approach to decision theory based on observations of decisions made during military exercises. During observations he studied the cognitive processes that occurred under conditions where decisions were made and responses were emitted in rapidly changing environments.

Klein (1993) emphasized the use of SMEs to determine appropriate decision standards and to evaluate the decision-making processes of less experienced decision makers. Because of the experience and knowledge SMEs possessed, evaluations made by them were considered to be more accurate than those made by less experienced decision makers. Several experiments (Baird, 1977; Borman & Hallam, 1991; Eubanks, Beal, & Bolla, 1998) from the industrial/organizational literature showed that reports of performance produced by experienced supervisors were more accurate than those provided by less experienced self-raters. However, more recent analyses (Facteau & Craig, 2001; Maurer, Raju, & Collins, 1998), along with previous evaluations (Farh, Werbal, & Bedeian, 1988; Harris & Schaubroeck, 1988) suggested that appraisals from supervisors were comparable with those made by self-raters. The latter studies differed from the former ones in that they did not compare performance

7

evaluations with actual performance. The absence of direct comparisons limited the ability to produce quantitative indicators of rating discrepancies.

These findings are relevant to the ratings discrepancies between cadre and students that emerged during the junior leader courses. Based on Klein's studies and results from experiments conducted by Baird (1977) and Borman and Hallam (1991), it might be assumed that cadre ratings were more accurate than self-ratings. If this were truly the case, then improvements reported by cadre were a more accurate reflection of performance improvements made by the students. However, results from several other experiments (Facteau & Craig, 2001; Farh, Werbal, & Bedeian, 1988; Harris & Schaubroeck, 1988; Maurer, Raju, & Collins, 1998) contradict this assumption. As stated above, the latter analyses did not include comparisons of reports with actual performance measures, whereas the former did. Therefore, based on direct comparison, it is safe to assume that ratings from experienced decision-makers are more accurate than those from inexperienced ones. Tremble's (1992) findings support this assumption. Ratings of leader competencies from experienced observer/controllers were more conservative than those from less-experienced subordinates but correlated highly with ratings of unit performance observed during training.

Another explanation of the rating discrepancies is based on the level of awareness inexperienced students bring to a leader course. Students who rate themselves at the beginning of a course are not yet aware of what they do not know. Experienced cadre who have observed new students' performance in previous courses may recognize initially that new students lack the knowledge and skills provided by course training. During the course, students may gain a greater awareness of their deficiencies, whereas cadre who recognized these deficiencies initially may witness improvements over time. These differing and changing perspectives held by students and cadre can lead to discrepancies in performance ratings at the beginning of the course and at its conclusion. This would also explain why self-ratings on some items from junior leader course 2 decreased over time. Students in course 2 may have become aware of their performance deficits to a greater extent than students from leader course 1, which would have induced rating decreases. Regardless of which ratings were more accurate, discrepancies between ratings from two different sources need to be explained in further detail.

Standards used to assess decision-making during naturalistic observations were determined by SMEs who differed in terms of their experience and knowledge (see Klein, 1997). The result was lack of a consistent standard, which was criticized by decision theorists (e.g., Howell, 1997). Schrader and Steiner (1996) provided empirical evidence that different standards of comparison produce different ratings. They suggested that raters usually relied on one of three comparison standards: (a) internal (comparison with self), (b) relative (comparison with peers), and (c) absolute (comparison with some objective measure). They also showed that under conditions where a comparison standard was not defined, discrepancies between rating sources were greater than when a standard was defined, regardless of whether the defined standard was internal, relative, or absolute. On the other hand, when self- and

8

supervisor raters used a well-defined absolute standard to rate performance, discrepancies between rating sources decreased to a greater extent than discrepancies produced by internal and relative comparison standards.

During the junior leader courses, no comparison standard for rating the student decision-making attributes was delineated. Nor were explicit attempts made to determine if standards used by cadre were similar to those used by the students. Discrepancies between cadre and student ratings suggested that comparison standards were ambiguous and/or that cadre used a different standard than students did. Thus, it was possible for cadre to give lower ratings to the same performance that students gave higher ratings.

Comparison standards for performance can be specified by providing appropriate instructions prior to training observations. Student and cadre raters can be provided with information regarding which comparison standard to implement, how that standard is defined, and how each rating category is to be interpreted. For example, if performance ratings are to be based on a relative comparison with peers, then raters can be instructed to "Rate the individual according to his performance relative to his fellow peers at this time." Rating categories and scales can also be defined. For example, to rate "adaptability," Knapp et al. (2001) delineated low ratings as follows: "Has difficulty functioning effectively in new situations; does not adapt quickly to new environments, people, or equipment; is easily frustrated in situations that do not go as planned (p. A-3)." Moderate ratings were defined as follows: "Is able to function adequately in new situations; modifies behavior when faced with unexpected events or conditions; adapts fairly quickly to new people, situations, or equipment (p. A-3)." High ratings were based on the following definition: "Thinks and acts quickly in response to changes in the environment; often develops innovative and imaginative approaches to dealing with unexpected events; can effectively change plans when the situation requires it (p. A-3)." Delineating performance comparison standards and providing definitions of rating categories prior to observations can help to alleviate ambiguity. Lessening ambivalence should result in fewer rating discrepancies across different rating sources.

Conclusion

Many researchers employ the use of self-report methods to assess training effectiveness because these techniques are usually non-intrusive and relatively easy to use. However, despite their ease of use, the burden of interpreting self-report data rests with researchers who adopt such methods. Self-report metrics were utilized during the two junior leader courses reported here. The data collected by these metrics were used to assess changes in the decision-making attributes of students over time. When discrepancies between student and cadre ratings arose, attempts were made to explain them by an appeal to previous research. This led to the conclusion that ratings provided by experienced cadre were more accurate representations of student decision-making performance than ratings made by the students themselves.

With regard to the use of performance ratings in general, quantitative analyses of inaccuracies can only occur when self-reports are compared to objective performance measures. However, in training environments such as leader courses, objective data collection methods are often intrusive, time consuming, prohibited, or all three. The difficulty of defining appropriate performance measures of decision-making also presents a problem. Defining a performance comparison standard, delineating the meaning of rating categories and scales, eliminating ambiguity whenever possible, and recognizing that students gain greater awareness of their limitations over time will help to improve the accuracy with which ratings of performance are made.

References

Baird, L. S. (1977). Self and superior ratings of performance: As related to self-esteem and satisfaction with supervisor. *Academy of Management Journal, 20,* 291-300.

Borman, W. C., & Hallam, G. L. (1991). Observation accuracy for assessors of work-sample performance: Consistency across task and individual-differences correlates. *Journal of Applied Psychology, 76,* 11-18.

Critchfield, T. S. (1993). Signal detection properties of verbal self-reports. *Journal of the Experimental Analysis of Behavior, 60,* 495-514.

Critchfield, T. S. (1996). Self-reports about performance under time-pressure: Bias and discriminability. *The Psychological Record, 46,* 333-350.

Critchfield, T. S., Tucker, J. A., & Vuchinich, R. E. (1998). Self-report methods. In K. A. Lattal & M. Perone (Eds.), *Handbook of research methods in human operant behavior* (pp. 435-470). New York: Plenum Press.

Department of the Army. (2001). *Operations* (Field Manual 3-0). Washington, DC: Author.

Department of the Army. (1993). *Operations* (Field Manual 100-5). Washington, DC: Author.

dt ogilvie, & Fabian, F. H. (1999). Decision-making requirements for future organizational leaders: A creative action-based approach. In J. G. Hunt, G. E. Dodge, & L. Wong (Eds.), *Out-of-the-box leadership: Transforming the twenty-first-century Army and other top-performing organizations* (pp. 63-90). Stamford, CT: JAI Press.

Eubanks, J. L., Beal, S. A., & Bolla, J. T. (1998, May). *Behavioral reporting in organizations: A contingency by vantage point analysis.* Paper presented at the 24th Annual Convention for the Association for Behavior Analysis, International, Orlando, FL.

Facteau, J. D., & Craig, S. B. (2001). Are performance appraisal ratings from different rating sources comparable? *Journal of Applied Psychology, 86,* 215-227.

Farh, J. L., Werbal, J. D., & Bedeian, A. G. (1988). An empirical investigation of self-appraisal-based performance evaluation. *Personnel Psychology, 41,* 141-156.

Harris, M. M., & Schaubroeck, J. (1988). A meta-analysis of self vs. supervisor, self vs. peer, and peer vs. supervisor ratings. *Personnel Psychology, 41,* 43-62.

Howell, W. C. (1997). Progress, prospects, and problems in NDM: A global view. In G. A. Klein & C. E. Zsambok (Eds.), *Naturalistic decision making* (pp. 37-48). Mahwah, NJ: Erlbaum.

Klein, G. A. (1993). A recognition-primed (RPD) decision model of rapid decision making. In G. A. Klein, J. Orasanu, R. Calderwood, & C. E. Zsambok (Eds.), *Decision making in action: Models and methods* (pp. 107-121). Norwood, NJ: Ablex.

Klein, G. A. (1997). An overview of naturalistic decision making applications. In G. A. Klein & C. E. Zsambok (Eds.), *Naturalistic decision making* (pp. 49-60). Mahwah, NJ: Erlbaum.

Knapp, D. J., Burnfield, J. L., Sager, C. E., Waugh, G. W., Campbell, J. P., Reeve, C. L., Campbell, R. C., et al. (2001). *Development of predictor and criterion measures for the NCO21 research program.* (ARI Technical Report 1128). Alexandria, VA: U.S. Army Research Institute for the Behavioral and Social Sciences.

Mabe, P. A., & West, S. G. (1982). Validity of self-evaluation of ability: A review and meta-analysis. *Journal of Applied Psychology, 67,* 280-296.

Maurer, T. J., Raju, N. S., & Collins, W. C. (1998). Peer and subordinate performance appraisal measurement equivalence. *Journal of Applied Psychology, 83,* 693-702.

Schrader, B. W., & Steiner, D. D. (1996). Common comparison standards: An approach to improving agreement between self and supervisory performance ratings. *Journal of Applied Psychology, 81,* 813-820.

Tremble, T. R. (1992). *Relationships of leadership competence with leader and unit performance effectiveness.* (ARI Research Report 1625). Alexandria, VA: U.S. Army Research Institute for the Behavioral and Social Sciences.

APPENDIX A

Leader Course Survey

The overall question is: How do you rate yourself on the following attributes?

Circle your answer using a seven point scale where
      1 = much lower than average performance
      2 = lower than average performance
      3 = slightly lower than average performance
      4 = average performance
      5 = slightly higher than average performance
      6 = higher than average performance
      7 = much higher than average performance
      CR = cannot rate - only if you have no basis on which to make a rating

NAME (for research tracking purposes only) _____ Date _____

How do you rate yourself on the following attributes?

1 much lower than average
2 lower than average
3 slightly lower than average
4 average
5 slightly higher than average
6 higher than average
7 much higher than average
CR cannot rate

Attribute	1	2	3	4	5	6	7	CR
Attention to detail	1	2	3	4	5	6	7	CR
Creativity	1	2	3	4	5	6	7	CR
Initiative, self-motivation, persistence	1	2	3	4	5	6	7	CR
Knowledge of military tactics/expertise	1	2	3	4	5	6	7	CR
Listening	1	2	3	4	5	6	7	CR
Seek innovative solutions	1	2	3	4	5	6	7	CR
Situation-awareness, perception	1	2	3	4	5	6	7	CR
Sound judgment, logical reasoning	1	2	3	4	5	6	7	CR
Time management	1	2	3	4	5	6	7	CR

Table A-1
Statistics on Junior Leader Course 1 Self-Ratings Times 1 & 2

Attribute		Self Time 1		Self Time 2					
	N	M	SD	M	SD	F	df	p	η^2
Knowledge of Military Tactics/Expertise	159	4.54	1.23	4.74	1.20	5.86	1,158	0.017	0.036
Time Management	161	5.19	1.12	5.32	1.10	2.21	1,160	0.139	0.014
Seek Innovative Solutions	159	5.34	1.02	5.32	1.00	0.05	1,158	0.821	0
Attention to Detail	161	5.08	1.17	5.32	1.03	9.84	1,160	0.002	0.058
Creativity	160	5.33	1.19	5.24	1.04	1.04	1,159	0.310	0.006
Situational Awareness, Perception	161	5.41	0.93	5.51	0.97	1.79	1,160	0.183	0.011
Listening	161	5.58	1.03	5.57	1.06	0.04	1,160	0.839	0
Sound Judgment, Logical Reasoning	161	5.73	0.83	5.65	0.88	1.29	1,160	0.257	0.008
Initiative, Self-Motivation, Persistence	161	5.62	1.01	5.49	1.03	2.70	1,160	0.102	0.017

Table A-2
Statistics on Junior Leader Course 1 Cadre Ratings Times 1 & 2

Attribute		Cadre Time 1		Cadre Time 2					
	N	M	SD	M	SD	F	df	p	η^2
Knowledge of Military Tactics/Expertise	144	4.25	0.86	4.65	0.99	24.81	1,143	0.001	0.148
Time Management	158	4.29	0.80	4.61	0.91	18.76	1,157	0.001	0.107
Seek Innovative Solutions	146	4.49	0.88	4.77	1.00	10.35	1,145	0.01	0.067
Attention to Detail	162	4.31	0.85	4.75	1.03	34.48	1,161	0.001	0.176
Creativity	145	4.51	0.85	4.69	1.01	4.15	1,144	0.05	0.028
Situational Awareness, Perception	162	4.53	0.79	4.69	1.01	4.19	1,161	0.05	0.025
Listening	161	4.69	0.87	4.79	1.01	2.17	1,160	0.142	0.013
Sound Judgment, Logical Reasoning	152	4.55	0.82	4.76	0.98	6.63	1,151	0.05	0.042
Initiative, Self-Motivation, Persistence	161	4.68	0.93	4.95	1.13	13.70	1,160	0.001	0.079

Table A-3
Statistics on Junior Leader Course 1 Self-Ratings and Cadre Ratings Time 1

Attribute	Self Time 1			Cadre Time 1					
	N	M	SD	M	SD	F	df	p	η^2
Knowledge of Military Tactics/Expertise	144	4.62	1.20	4.25	0.86	12.39	1,143	0.001	0.080
Time Management	158	5.13	1.14	4.29	0.80	78.26	1,157	0.001	0.333
Seek Innovative Solutions	145	5.28	1.00	4.50	0.88	53.58	1,144	0.001	0.271
Attention to Detail	162	5.05	1.16	4.31	0.85	45.68	1,161	0.001	0.221
Creativity	145	5.30	1.17	4.51	0.85	45.02	1,144	0.001	0.238
Situational Awareness, Perception	162	5.39	0.95	4.53	0.79	98.35	1,161	0.001	0.379
Listening	161	5.55	1.04	4.69	0.87	69.17	1,160	0.001	0.302
Sound Judgment, Logical Reasoning	152	5.68	0.83	4.55	0.82	161.60	1,151	0.001	0.517
Initiative, Self-Motivation, Persistence	160	5.59	1.03	4.68	0.93	80.60	1,159	0.001	0.336

Table A-4
Statistics on Junior Leader Course 1 Self-Ratings and Cadre Ratings Time 2

Attribute	Self Time 2			Cadre Time 2					
	N	M	SD	M	SD	F	df	p	η^2
Knowledge of Military Tactics/Expertise	133	4.78	1.21	4.65	0.98	1.45	1,132	0.239	0.011
Time Management	148	5.28	1.11	4.61	0.92	37.99	1,147	0.001	0.205
Seek Innovative Solutions	135	5.30	1.00	4.78	0.98	19.92	1,134	0.001	0.129
Attention to Detail	152	5.30	1.04	4.75	1.02	26.25	1,151	0.001	0.148
Creativity	135	5.23	1.04	4.70	0.98	16.73	1,134	0.001	0.111
Situational Awareness, Perception	152	5.47	0.97	4.68	0.98	59.79	1,151	0.001	0.284
Listening	151	5.54	1.07	4.78	0.99	43.46	1,150	0.001	0.225
Sound Judgment, Logical Reasoning	142	5.63	0.84	4.76	0.96	75.61	1,141	0.001	0.349
Initiative, Self-Motivation, Persistence	151	5.46	1.04	4.96	1.12	19.61	1,150	0.001	0.116

Table A-5
Statistics on Junior Leader Course 2 Self-Ratings Times 1 & 2

Attribute	N	Self Time 1		Self Time 2					
		M	SD	M	SD	F	df	p	η^2
Knowledge of Military Tactics/Expertise	132	4.59	1.17	4.63	1.19	0.20	1,131	0.654	0.002
Time Management	131	5.02	1.14	5.04	1.13	0.07	1,130	0.791	0.001
Seek Innovative Solutions	131	5.15	1.01	5.10	0.93	0.37	1,130	0.546	0.003
Attention to Detail	132	5.17	1.06	5.10	0.98	5.81	1,131	0.447	0.004
Creativity	132	5.28	1.04	5.04	1.11	8.80	1,131	0.004	0.063
Situational Awareness, Perception	131	5.31	0.98	5.22	0.98	1.30	1,130	0.256	0.01
Listening	131	5.47	0.97	5.24	0.95	6.49	1,130	.012	0.048
Sound Judgment, Logical Reasoning	131	5.51	0.91	5.34	0.91	5.41	1,130	.022	0.04
Initiative, Self-Motivation, Persistence	132	5.57	1.01	5.30	0.98	10.19	1,131	0.002	0.072

Table A-6
Statistics on Junior Leader Course 2 Cadre Ratings Times 1 & 2

Attribute	N	Cadre Time 1		Cadre Time 2					
		M	SD	M	SD	F	df	p	η^2
Knowledge of Military Tactics/Expertise	69	4.22	1.06	4.83	1.10	22.80	1,68	0.001	0.251
Time Management	76	4.32	0.94	4.89	1.00	30.44	1,75	0.001	0.289
Seek Innovative Solutions	74	4.29	0.87	5.02	0.93	43.52	1,73	0.001	0.374
Attention to Detail	76	4.44	0.90	4.99	1.03	28.29	1,75	0.001	0.274
Creativity	74	4.24	0.93	4.96	0.94	34.83	1,73	0.001	0.323
Situational Awareness, Perception	77	4.45	1.07	5.03	0.98	24.39	1,76	0.001	0.243
Listening	77	4.58	1.04	4.97	1.19	8.99	1,76	0.01	0.106
Sound Judgment, Logical Reasoning	76	4.43	0.95	5.06	0.93	29.47	1,75	0.001	0.282
Initiative, Self-Motivation, Persistence	76	4.47	0.95	5.09	1.11	23.37	1,75	0.001	0.238

Table A-7
Statistics on Junior Leader Course 2 Self-Ratings and Cadre Ratings Time 1

Attribute	N	Self Time 1		Cadre Time 1					
	N	M	SD	M	SD	F	df	p	η^2
Knowledge of Military Tactics/Expertise	68	4.22	1.20	4.22	1.06	0.002	1,67	0.967	0
Time Management	76	4.93	1.22	4.32	0.90	13.83	1,75	0.001	0.156
Seek Innovative Solutions	74	5.23	1.01	4.29	0.87	42.94	1,73	0.001	0.370
Attention to Detail	76	5.24	0.99	4.44	0.90	27.14	1,75	0.001	0.266
Creativity	74	5.24	1.10	4.24	0.93	36.08	1,73	0.001	0.331
Situational Awareness, Perception	77	5.29	1.07	4.46	1.08	29.02	1,76	0.001	0.276
Listening	77	5.30	1.05	4.58	1.04	17.06	1,76	0.001	0.183
Sound Judgment, Logical Reasoning	77	5.54	0.96	4.43	0.95	59.18	1,76	0.001	0.441
Initiative, Self-Motivation, Persistence	75	5.61	1.01	4.49	0.94	51.60	1,74	0.001	0.411

Table A-8
Statistics on Junior Leader Course 2 Self-Ratings and Cadre Ratings Time 2

Attribute	N	Self Time 2		Cadre Time 2					
	N	M	SD	M	SD	F	df	p	η^2
Knowledge of Military Tactics/Expertise	69	4.54	1.20	4.83	1.10	2.34	68	0.130	0.033
Time Management	75	4.93	1.22	4.88	1.00	0.100	74	0.753	0.001
Seek Innovative Solutions	73	5.15	1.06	5.03	0.93	0.48	72	0.493	0.007
Attention to Detail	76	5.05	1.22	4.99	1.03	0.12	75	0.729	0.002
Creativity	75	5.35	1.16	4.95	0.94	4.88	74	0.030	0.062
Situational Awareness, Perception	76	5.33	1.00	5.01	0.98	4.12	75	0.046	0.052
Listening	77	5.44	0.99	4.97	1.19	6.41	76	0.013	0.078
Sound Judgment, Logical Reasoning	75	5.52	0.95	5.05	0.93	8.70	74	0.004	0.105
Initiative, Self-Motivation, Persistence	77	5.40	1.23	5.11	1.12	2.24	76	0.139	0.029

Table A-9
Percentage of Student Responses and Changes Over Time

Attribute	JLC 1 Self-Ratings			JLC 2 Self-Ratings		
	No Change	Worse Rating	Improved	No Change	Worse Rating	Improved
Knowledge of Military Tactics/Expertise	42	21	37	38	31	31
Time Management	39	28	33	42	28	30
Seek Innovative Solutions	46	27	27	51	27	22
Attention to Detail	43	19	38	46	30	24
Creativity	47	29	24	43	37	20
Situational Awareness, Perception	46	24	30	52	27	21
Listening	40	28	32	41	39	20
Sound Judgment, Logical Reasoning	50	29	21	47	34	19
Initiative, Self-Motivation, Persistence	42	33	25	45	37	18

Table A-10
Decision-Making Items Correlations

Attribute	JLC 1 Self Times 1 & 2	JLC 1 Cadre Times 1 & 2	JLC 2 Self Times 1 & 2	JLC 2 Cadre Times 1 & 2
Knowledge of Military Tactics/Expertise	0.63	0.44	0.66	0.51
Time Management	0.50	0.42	0.62	0.57
Seek Innovative Solutions	0.46	0.41	0.60	0.44
Attention to Detail	0.61	0.52	0.50	0.47
Creativity	0.58	0.33	0.60	0.37
Situational Awareness, Perception	0.51	0.43	0.63	0.51
Listening	0.38	0.57	0.46	0.47
Sound Judgment, Logical Reasoning	0.36	0.40	0.55	0.43
Initiative, Self-Motivation, Persistence	0.51	0.60	0.52	0.43

Note: All correlations significant $p < .01$.